Detecting Original Image Using Histogram, DFT and SVM

T. H. Manjula Devi[1], H.S.Manjunatha Reddy,[2] K. B. Raja[3]Venugopal K. R[3] and L. M. Patnaik[4]
[1]Department of Telecommunication Engg, Dayananda Sagar College of Engineering, Bangalore
[2]Department of Electronics & Communication, Global Academy of Technology
[3]Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore
[4]Vice chancellor, Defence Institute of Advanced Technology, Puna.
thm0303@gmail.com

Abstract

Information hiding for covert communication is rapidly gaining momentum. With sophisticated techniques being developed in steganography, steganalysis needs to be universal. In this paper we propose Universal Steganalysis using Histogram, Discrete Fourier Transform and SVM (SHDFT). The stego image has irregular statistical characteristics as compare to cover image. Using Histogram and DFT, the statistical features are generated to train One-Class SVM to discriminate the cover and stego image. SHDFT algorithm is found to be efficient and fast since the number of statistical features is less compared to the existing algorithm.

Keywords: Universal Steganalysis, Cover-Medium, Payload, Redundant Bits, Total Cover Capacity, Histogram, DFT, SVM.

I. INTRODUCTION

Since long time ago, surreptitious communication and exchange of information is well known. There are numerous illustrations depicting covertness used in communication. Steganography, water marking, cryptography are some of the techniques used for hiding information [1]. There are several mediums for hiding information, such as digital images, text, audio, web pages, video, etc. Steganography is a skill and discipline of embedding a secret payload into a cover medium. The redundant bits in the cover medium are identified and replaced with the payload. The intent of steganography is ensuring that the presence of hidden message is undetectable. The different techniques used for embedding of data are *Least Significant Bit (LSB), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Spread Spectrum (SS) and Palette technique*. There are software tools available over the Internet for embedding the payload in digital images and videos viz., S-Tools, J-Steg, Outguess, F5 and Steghide.

The steganography is being used by criminals and terrorists to exchange information regarding their illegal activities. The government and other standard organizations are using steganalysis to restrain these illegal activities. Steganalysis is to discover and recognize the extent of a hidden message and can be either blind (universal) or non-blind (embedding specific). In case of universal technique, the embedding scheme is unknown; so attempt is made to detect the existence of hidden data. Whereas in embedding specific, the embedding technique is known; therefore payload size is estimated. There are Steganalysis Tools available freely or as commercial software. The Stegdetect Tool, detects *Jsteg, Jphide, invisible secrets, Outguess, F5, camouflage* Steganographic schemes in JPEG images. The Tools developed using the Chi-Square analysis, performs a statistical attack to detect hidden data in BMP images.

The success of steganography depends on various aspects such as the algorithm used for embedding, compression algorithm used and alteration of image properties. LSB is most commonly used technique for image steganography, which uses bitwise methods to manipulate LSB of the cover image. The minute changes of LSB are imperceptible to human eye and the method of hiding information in LSB can be analogized by adding noise to the image. Reliable algorithms for compression used in Steganography are Windows Bitmap (BMP), Graphics Interchange Format (GIF) and Joint Photographic Experts Group (JPEG), to ensure the hidden information is not lost after transformation. Lossless compression algorithms such as BMP and GIF formats are chosen for LSB techniques in which modifications are usually made in spatial domain. JPEG is losy compression algorithm where data hiding is usually done in frequency domain.

General Steganalysis method has not been developed since every Steganographic method utilizes different methods of embedding the payload. In classical Steganographic schemes, the security lies in the concealment of the encoding technique whereas the modern schemes adopt Kerchoff's Principle of Cryptography, and hence the security depends on the secret key that is used to encode the payload.

There are several Classifiers being used for pattern recognition: *Bayesian Multi-Variate, Fischer Linear Discriminant (FLD), Neural Network (NN), and Support Vector Machines (SVM).* SVM is based on statistical learning theory which is highly dependent upon the kernel functions viz., Gaussian Radial Basis Function, Polynomial, Exponential Radial Basis Function, Multi-Layer Perceptron, Splines, BSplines, Additive Kernels and Tensor Product are used for mapping features. SVM can be classified as One-Class and Multi-Class. In One-

DOI: 01.ijsip.01.01.04

Class SVM, only the features of cover images are required for classification of cover and stego image whereas in Multi-Class SVM, the features of both the cover and stego-images are required for discriminating the cover and stego image. Owing to simple geometry, SVMs are used more predominantly than Neural Networks. Biggest limitation of SVM is choosing of the kernel function, speed and size, both in training and testing.

Contribution: In this paper, we have proposed SHDFT Universal Steganalysis Method. Histogram and Discrete Fourier Transform of color images are computed to obtain the PDF moments. In addition, the energy of the image is considered. The non-linear Classifier Support Vector Machine is used for classification of cover and stego image.

II. RELATED WORK

Anderson and Petitcolas [2] have proposed two mathematical frame works for Steganography: informatics and theoretical models. The quality of the retrieved image is poor in this algorithm. Petitcolas et al., [3] have discussed a survey on information hiding. The applications and the terminologies used in information hiding are elucidated. Certain limitations of information hiding have been portrayed, the JPEG compression, low pass filtering, cropping, scaling, Additive Gaussian noise, rotation, cause distortion in the image. An important understanding of these limitations is helpful in steganalysis.

Fridich et al., [4] have presented a steganalytic method to identify stego image generated by F5 steganographic algorithm in JPEG images reliably. The method involves estimation of the cover image histogram from the stego image. The statistical variation of cover image is determined using the Least Square Fit, which compares estimated histogram of DCT coefficients of stego-image. This technique can be extended to other steganographic algorithms that manipulate quantized DCT coefficients. Siwei Lyu and Hany Farid [5] have described an approach of multi-scale Wavelet Decomposition to build Higher Order Statistical Model to detect hidden data in gray images. Support Vector Machine is used to detect the statistical variations in a test image.

Farid [6] has described a steganalytic method to detect hidden messages using Wavelet Decomposition by building Higher Order Statistical Model of the image. The model extracts basic coefficient statistics and error statistics from an optimal magnitude predictor. FLD is used to distinguish the cover and stego image. Siwei Lyu and Hany Farid [7] have described a Universal Steganalysis Algorithm that exploits strong statistical regularities that exist between the color channels of natural images. The statistical model is generated using first and Higher Order Statistics of the image. One Class SVM is used to detect hidden data.

Hany Farid and Siwei Lyu [8] have employed Multi-Scale Wavelet Decomposition to build a statistical model for images. The statistical regularities exhibited by the natural images are captured by First and Higher Order Statistical Model. The FLD is used for classification of the cover and stego image. This model explores its usefulness in digital forensics including steganography; distinguish between natural photographs and computer generated images.

Siwei Lyu and Hany Farid [9] have proposed universal Steganalysis using Higher Order Statistics (SHS). Wavelet Decomposition based on QMF is used to obtain the statistics of an image. The magnitude statistics is obtained from Linear Predictor and phase statistics is obtained from Gaussian Pyramid and Local Angular Harmonic Decomposition. One Class and Multi Class SVM is used for classification. The disadvantage of the algorithm is the number of features required to train the SVM are high. Hence, the time required to detect stego image is more.

III. MODEL

The steganalysis model is discussed in this section.

Steganalysis Model

Figure 1 gives the block diagram of SHDFT to discriminate cover and stego images using SVM.

Color image: JPEG and BMP color images are used for the analysis. It consists of RGB color planes, $c \in \{r, g, b\}$. The color planes are separated and 1D-Histogram is applied to each of the separated color planes of the image.

Histogram: Histograms are the basis for numerous spatial domain processing techniques used to provide useful image statistics. Image processing operations such as steganography result in changes to the image histogram. The histogram is applied on each color plane of the image. The first and second *PDF* moments i.e., the mean and the variance of the histogram coefficients of each colour plane are calculated which yields *6D* statistical features.

Discrete Fourier Transform (DFT): The sequence of N spatial complex coefficients $x_0, x_1, \ldots\ldots, x_{N-1}$ is transformed into the sequence of N frequency complex coefficients $X(0), X(1), \ldots\ldots, X(N-1)$ by the DFT according to the formula as given in Equation

$$X(k) = \sum_{n=0}^{N-1} x(n)\exp\left(\frac{-i2\pi nk}{N}\right)$$

Where n=0, 1… N-1 and k=0, 1… N-1.

DFT for Histogram of each color channel, $c \in \{r, g, b\}$ is applied. The *PDF* moments *i.e., mean, variance, skewness* and *kurtosis* of the *DFT* coefficients are computed which yields *12D* statistics. The total energy for DFT coefficients is computed which yields 3D Statistics.

SVM: In One-Class SVM, only the features of cover images are required for classification of cover and stego image. The feature vector comprising *24D* statistics is used to train the SVM.

DOI: 01.ijsip.01.01.04

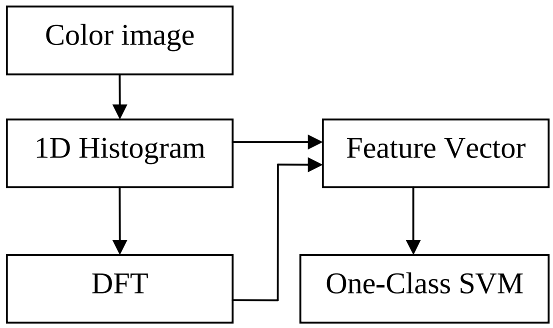

Figure 1: Block Diagram of SHDFT

Features: We find the mean of the difference between *DFT* and *Histogram* for each color plane which yields *3D* statistics. SHDFT totally yields *24D* characteristics which are used to discriminate the cover and stego image.

IV. ALGORITHM

TABLE 1:

ALGORITHM FOR SHDFT

Algorithm for SHDFT

- **Input**: The test image
- **Output**: SVM classifies the test image as cover or stego image.

1. Separation of the color planes, $c \in \{r, g, b\}$ of the color image.
2. Build Histogram (*1D*) for each color planes $c \in \{r, g, b\}$. Compute the 1st and 2nd moments i.e., mean and variance of the histogram coefficients. This yields *6D* statistics.
3. Build Discrete Fourier Transform for histogram of each color planes, $c \in \{r, g, b\}$. Compute the *1st, 2nd, 3rd and 4th* moments i.e., mean variance, skewness and kurtosis of the DFT coefficients. This yields *12D* statistics. Also compute the total energy for the DFT coefficients. This yields *3D* more statistics.
4. The *1st* order moment Mean of the difference of histogram and DFT is computed for each color channel. This yields *3D* statistics.
5. The feature vectors obtained from steps *2, 3* and *4* yields *24D* features.
6. The feature vectors obtained in steps 5 are used to train SVM to classify test image.

V. PERFORMANCE ANALYSIS

For the performance analysis we consider 100 color images of JPEG and BMP formats with 600*400 pixels in size (86 kilobytes). Figure 2 (a)(b) shows the payload. Figure2 (c)(d) and (e)(f) gives the cover images of BMP and JPEG formats respectively. The payload of various sizes is embedded into the full-resolution cover images to generate 100 stego images Using Jphide software for JPEG images and S-Tools 4.0 software for BMP images. The payload is of the size 6.0, 4.7, 1.2, 0.3 kilobytes, corresponding to 100%, 78%, 20% and 5% Of total cover capacity The capacity will vary according to the size of the cover image. The steganographic capacity is then the ratio of the size of the embedded payload to the total cover capacity. Each stego image is generated with the same quality factor as that of the original cover image so as to minimize the statistical variations. Figure 3 and Table 2 gives the comparison of the detection rate for the proposed algorithm SHDFT and existing algorithm SHS for different embedding rate in BMP images using S-Tools. It is observed that the percentage of detection rate increases as embedding rate increases for both the algorithms. Figure 4 and Table 3 gives the comparison of the detection rate for the proposed algorithm SHDFT and existing algorithm SHS for different embedding rate in JPEG images using *JPhide* and *S-Tools*. The proposed algorithm SHDFT requires only *24D* statistics for training SVM as compared *to 432D* statistics required for existing SHS algorithm for the detection of payload[9].

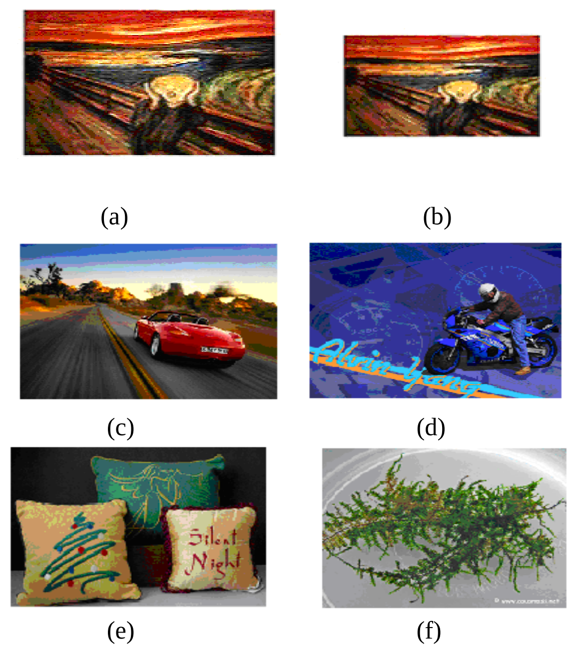

Figure 2: The payload images (a)(b), BMP cover image (c)(d) and JPEG cover image (e)(f).

DOI: 01.ijsip.01.01.04

TABLE 2

EMBEDDING RATE FOR BMP IMAGES USING S-TOOLS

Embedding Rate (%)	Detection Rate (%)	
	SHDFT	SHS
100	96.42	85.71
78	92.67	78.57
20	89.87	71.42
5	86.78	64.28

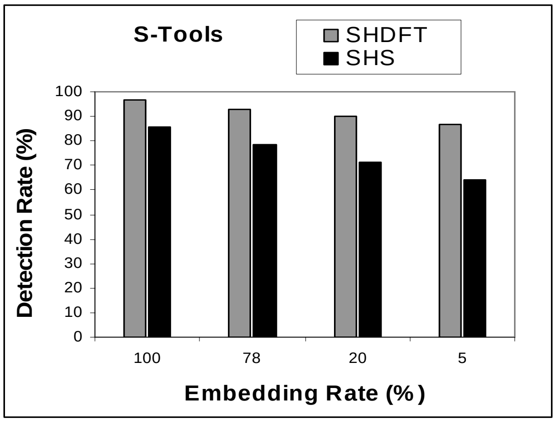

Figure 3: Comparison of Detection rate for various payload sizes for SHDFT and SHS algorithm for BMP images using OC-SVM.

TABLE 3
EMBEDDING RATE FOR JPEG IMAGES USING JPHIDE

Embedding Rate (%)	Detection Rate (%)	
	SHDFT	SHS
100	98.67	89.91
78	95.73	80.57
20	91.32	76.42
5	86.78	68.28

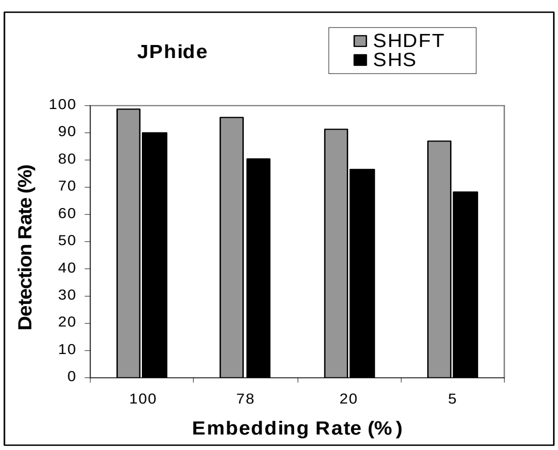

Figure 4: Comparison of Detection rate for various payload sizes for SHDFT and SHS algorithm for JPEG images using OC-SVM

VI. CONCLUSION

In the wake of a number of steganographic techniques being developed, steganalysis needs to have a universal approach. We present a steganalysis technique that relies on building a Statistical Model using the Histogram and Discrete Fourier Transform. SVM is used for classification of stego and cover images. This Model has fairly higher detection rates and comparatively computation time is less, as the numbers of features required to train the SVM is much lesser than other techniques. In future, the algorithm may be tested for embedding rate less than 5%.

REFERENCES

[1] D. Kahn, "The History of Steganography," *Lecture Notes in Computer Science*, Springer-Verlag, vol. 1174, 1996.

[2] R. Anderson and F. Petitcolas, "On the Limits of Steganography," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 4, pp. 474–481, 1998.

[3] E. Petitcolas, R. Anderson and M. Kuhn, "Information Hiding – a Survey," *Proceedings of the IEEEs*, vol. 87, no. 7, pp. 1062–1078, 1999.

[4] J. Fridrich, M. Goljan, and D. Hogea, "Steganalysis of JPEG images: Breaking the F5 Algorithm," *International Workshop on Information Hiding, Lecture Notes in Computer Science*, Springer-Verlag, vol. 2578, pp. 310-323, 2003.

[5] S. Lyu and H. Farid, "Detecting Hidden Messages using Higher Order Statistics and Support Vector Machines,", *International Workshop on Information Hiding, Lecture Notes in Computer Science*, Springer-Verlag, vol. 2578, pp. 340-354,

[6] S. Lyu and H. Farid, "Steganalysis using Color Wavelet Statistics and One-Class Support Vector Machines," in *SPIE Symposium on Electronic Imaging*, vol. 5306, pp. 35-46, 2004.

[7] E. P. Simoncelli, "Statistical Modelling of Photographic Images," in *Handbook of Image and Video Processing*, A. Bovik, Ed. Academic Press, chapter 4.7, Second Edition, 2005.

[8] Hany Farid and Siwei Lyu, "Higher-order Wavelet Statistics and their Application to Digital Forensics," *IEEE Workshop on Statistical Analysis in Computer Vision*, 2003.

[9] Siwei Lyu and Hany Farid, "Steganalysis using Higher-Order Image Statistics," *IEEE Transaction on Information Forensics and Security*, vol. 1, pp. 111-119, 2006.

Mrs. Manjula Devi T.H is an Assistant Professor in the Department of Telecommunication Engg, Dayananda Sagar College of Engg, Bangalore. She obtained her B.E. degree in Electronics Engg from Bangalore University. Her specialization in Master degree was Electronics & Communication from

DOI: 01.ijsip.01.01.04

Bangalore University and currently she is pursuing Ph.D. in the area of Steganography and Steganalysis under the guidance of Dr. K. B. Raja, Asst. Professor, Dept of Electronics and Communication Engg, University Visvesvaraya college of Engg, Bangalore. Her area of interest is in the field of Signal Processing, Communication and Control Systems. She is the life member of International Microelectronics And Packaging Society (IMAPS) India Chapter, and Indian society for Technical Education.

Mr. H.S.Manjunatha Reddy is a Professor in the department of Electronics and Communication Engineering, Global Academy of Technology, Bangalore. He obtained his B.E. Degree in Electronics from Bangalore University, Bangalore. His specialization in Master degree was Digital Electronics from Visvesvaraya Technological University, Belgaum. He is pursuing research in the area of Steganography and Steganalysis for secured communication. His area of interest is in the field of Digital Image Processing, Communication Networks and Biometrics. He is life member of Indian Society for Technical Education, New Delhi.

K B Raja is an Assistant Professor, Dept. of Electronics and Communication Engg, University Visvesvaraya college of Engg, Bangalore University, Bangalore. He obtained his Bachelor of Engineering and Master of Engineering in Electronics and Communication Engineering from University Visvesvaraya College of Engineering, Bangalore. He was awarded Ph.D. in Computer Science and Engineering from Bangalore University. He has over 35 research publications in refereed International Journals and Conference Proceedings. His research interests include Image Processing, Biometrics, VLSI Signal Processing, computer networks.

K R Venugopal is currently the Principal and Dean, Faculty of Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bangalore. He obtained his Bachelor of Engineering from University Visvesvaraya College of Engineering. He received his Masters degree in Computer Science and Automation from Indian Institute of Science Bangalore. He was awarded Ph.D. in Economics from Bangalore University and Ph.D. in Computer Science from Indian Institute of Technology, Madras. He has a distinguished academic career and has degrees in Electronics, Economics, Law, Business Finance, Public Relations, Communications, Industrial Relations, Computer Science and Journalism. He has authored 27 books on Computer Science and Economics, which include Petrodollar and the World Economy, C Aptitude, Mastering C, Microprocessor Programming, Mastering C++ etc. He has been serving as the Professor and Chairman, Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bangalore. During his three decades of service at UVCE he has over 200 research papers to his credit. His research interests include computer networks, parallel and distributed systems, digital signal processing and data mining.

L M Patnaik is a Vice Chancellor, Defence Institute of Advanced Technology (Deemed University), Pune, India. During the past 35 years of his service at the Indian Institute of Science, Bangalore. He has over 500 research publications in refereed International Journals and Conference Proceedings. He is a Fellow of all the four leading Science and Engineering Academies in India; Fellow of the IEEE and the Academy of Science for the Developing World. He has received twenty national and international awards; notable among them is the IEEE Technical Achievement Award for his significant contributions to high performance computing and soft computing. His areas of research interest have been parallel and distributed computing, mobile computing, CAD for VLSI circuits, soft computing, and computational neuroscience.

DOI: 01.ijsip.01.01.04